

Remarks

The Present Invention and the Pending Claims

The present invention relates generally to the field of speech recognition. More particularly, the invention discloses a technique for disambiguating speech input using one of voice mode interaction, visual mode interaction, or a combination of voice mode and visual mode interaction.

selecting two or more tokens generated without translation of the language in which the speech, audio or combination of speech and audio input is received, and presenting said tokens as alternatives to be presented to the user as alternatives, wherein said alternatives are words or tokens;

presenting the alternatives to the user in one of voice mode, visual mode, or a combination of voice and visual mode, and receiving a selection of an alternative from by the user from the plurality of alternatives presented to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode; and,

communicating the selected alternative to the application as input to the application.

Claim 12 (original). The method of claim 11, wherein the interaction comprises the concurrent use of said visual mode and said voice mode.

Claim 13 (original). The method of claim 12, wherein the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input.

Claim 14 (original). The method of claim 11, wherein the interaction comprises the user selecting from among the plural alternatives using visual input.

Remarks

The Present Invention and the Pending Claims

The present invention relates generally to the field of speech recognition. More particularly, the invention discloses a technique for disambiguating speech input using one of voice mode interaction, visual mode interaction, or a combination of voice mode and visual mode interaction.

Claims 1, 4-5, 7-8, and 11-14 are currently pending. Reconsideration and allowance of the pending claims is respectfully requested.

Summary of the Office Action

Claims 1, 4, 7-8, 11 and 14 rejected under 35 U.S.C. 103(a) by Lai et al. (USPN 6,006,183) referred to as Lai hereinafter in view of Duan et al. (US Patent No. 6,223,150) referred to as Duan hereinafter.

Claims 5 and 12-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lai in view of Haddock et al. (USPN 5,265,014) referred to as Haddock hereinafter.

Amendments To The Claims

Claims 1 and 11 are currently amended. Support for the amendments in claims 1 and 11 are found at paragraphs [0017], [0021], [0022], [0024], [0025], [0027], and [0028].

The office action states: "Claims 1, 4, 7-8, 11 and 14 rejected under 35 U.S.C. 103(a) by Lai et al. (USPN 6,006,183) referred to as Lai hereinafter in view of Duan et al. (US Patent No. 6,223,150) referred to as Duan hereinafter."

First, Lai in view of Duan, does not teach or suggest all the claim limitations. The applicant's invention discloses an options and parameters 114 component for setting

6

"presenting the alternatives to the user in one of voice mode, visual mode, or a combination of voice and visual mode, and receiving a selection of an alternative by the user from the plurality of alternatives presented to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode" in claim 11.

Therefore, Lai in view of Duan does not teach or suggest the limitations in amended claims 1 and 11.

Furthermore, even if Lai and Duan are combined as suggested in the office action, the combination that results will not arrive at the claimed invention. For example, Lai in view of Duan will be inoperable and unsuccessful for the purpose of arriving at the system or corresponding method steps in claim 1 and 11 respectively, because Lai in view of Dunn does not teach:

- "an application parameter for controlling the speech disambiguation mechanism",
- "a speech recognition component that .. generates one or more tokens corresponding to the speech input without translation of the language in which the audio, speech, or a combination of speech and audio input is received and presents said tokens as alternatives to the user",
- "one or more disambiguation components that present the alternatives to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode", and
- "receives an alternative selected by the user from the plurality of alternatives presented to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode".

Furthermore, applicant respectfully submits that the Lai and Duan references that are sought to be combined are in non-analogous arts. Applicant's invention is a speech recognition system where tokens are generated based on recognition of the speech uttered by the user in a single language by a speech recognizer (see paragraph [0011]). In

9

resented to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode;".

In summary, Lai in view of Duan does not teach the following limitations in claim 1 and 11:

"an options and parameters component for receiving user parameters and application parameters for controlling the speech disambiguation mechanism,

parameters required for controlling the disambiguation mechanism by the user and the application (see paragraphs [0017], [0021], [0022], [0024]; [0025]; and Fig 1; paragraph [0024] recites: "The end user 108 and the application 106 can both set parameters 114 to control the sub-components of the MDM"). The limitation: "receiving parameters from a user and the application for controlling the speech disambiguation mechanism, wherein both the user and the application can set the parameters to control said mechanism..." in claim 1 and 11 is not found in Lai or Duan. In contrast to the setting of parameters by both the user and the application to control the speech disambiguation mechanism in applicant's invention, neither Lai or Duan teach setting of parameters by the application. Lai teaches assigning a confidence level score by a confidence level score 200 of the speech engine 160 (Lai, col. 3, lines 29-30); enabling a user of the system to select score thresholds (Lai, col. 3, lines 37-42); and allowing the user application to accept information from the user control (Lai, col. 4, lines 11-15).

contrast, Duan is a language translation system where tokens are generated based on translation of a user's utterance from a source language to a target language (see Duan, col. 21, lines 9-13; and col. 9, lines 55-65). The method by which tokens are generated by a speech recognition system for speech uttered in a single language is vastly different and distinguishable from the method in Duan where tokens are generated by a language translation system that translates a user's utterance from a source language to a target language. A person of ordinary skill in the art would not likely look at a language translation system to find how to find how a plurality of alternative words can be generated and presented to a user, for use in a speech recognition system for the reason stated above. Therefore, applicant respectfully submits that the teachings of Lai and Duan may not be combined.

Claims 4, 7-8 are dependent on claim 1 and further claim 14 is dependent on claim 11. Since Lai and Duan does not teach, suggest, or motivate the limitations of claim 1 and 11, applicant respectfully submits claims 4, 7-8, and 14 also to be novel over Lai and Duan. The applicant solicits reconsideration and allowance of claims 1, 4, 7-8, 11, and 14.

The office action states: "Claims 5 and 12-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lai in view of Haddock et al. (USPN 5,265,014) referred to as Haddock hereinafter".

Even if Lai and Haddock are combined as suggested by the Examiner, the combination that results will be inoperable for the purpose intended by claim 1 in applicant's invention, i.e., "an options and parameters component for receiving user parameters and application parameters for controlling the speech disambiguation mechanism, wherein both the user and the application can set the parameters to control said mechanism, and wherein the parameters include confidence thresholds governing unambiguous recognition and close matches". In applicant's invention, the parameters are set by both the user and the application. In contrast, neither Lai and Haddock teach or suggest setting of the parameters by both the user and the application.

resented to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode;".

In summary, Lai in view of Duan does not teach the following limitations in claim 1 and 11:

"an options and parameters component for receiving user parameters and application parameters for controlling the speech disambiguation mechanism, wherein both the user and the application can set the parameters to control said mechanism, ..." in claim 1, and

"a selection component that identifies, according to a selection algorithm, which two or more tokens generated without translation of the language in which the speech, audio or combination of speech and audio input is received, and presents said tokens as alternatives to the user" in claim 1, and

"one or more disambiguation components that present the alternatives to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode, and receive an alternative selected by the user from the plurality of alternatives presented to the user in one of voice mode, visual mode, or a combination of voice mode and visual mode" in claim 1, and

"receiving user parameters and application parameters for controlling the speech disambiguation mechanism, wherein both the user and the application can set the parameters to control said mechanism, and wherein the parameters include confidence thresholds governing unambiguous recognition and close matches" in claim 11

"selecting two or more tokens generated without translation of the language in which the speech, audio or combination of speech and audio input is received, and presenting said tokens as alternatives to the user" in claim 11, and

The applicant's system for disambiguating speech input comprises, in part, of a disambiguation component that presents two or more alternatives to a user in voice mode, visual mode, or a combination of voice mode and visual mode, and receives an alternative selected by the user in voice mode, visual mode, or a combination of voice mode and visual mode (see paragraph [0028]). However, neither Lai nor Haddock teach selection of the alternatives by the user using a combination of voice mode and visual mode. As stated on page 2 and 3, items i. to v. of the office action and Figure 2, Lai suggests that an acoustic signal may be inputted to the speech recognizer 190 and the system may display words with confidence level indicated. However, Lai does not expressly teach that the preferences are provided to the user to select in voice mode or visual mode or a combination of voice mode and visual mode, but instead suggests that the input to the speech recognizer 190 is in voice mode and preferences are provided to the user in visual mode. Hence, there is no teaching, suggestion or motivation in Lai and Haddock of the following limitations recited in applicant's claims 5, 12 and 13:

"wherein the disambiguation components present the alternatives to the user in a visual form and allow the user to select from among the alternatives using a voice input" of claim 5,

"where the interaction comprises the concurrent use of said visual mode and said voice mode" of claim 12, and

"wherein the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input" of claim 13.

The disambiguation components such as the output generator and the input handler in the present invention (see paragraphs [0027] and [0028]) allow multimodal interaction including voice mode interaction, visual mode interaction, or a combination of voice mode and visual mode for a user with the disambiguation mechanism for the purposes of disambiguating speech input in case of an ambiguous speech input recognition. Claims 5, 12, and 13 (by virtue of their dependence on claims 1 and 11)

recite different instances of multimodal interaction of the user with the system for disambiguating speech input using the disambiguation components.

Common sense dictates that a person of ordinary skill in the art, at the time the invention was made, would not combine the method of indicating the level of confidence the system has in its speech recognition as described in Lai and the method for disambiguating natural language queries using referential input by a user as described in Haddock, to arrive at the claimed invention because Lai and/or Haddock show no recognition or appreciation of the following limitations recited in claims 5, 12, and 13:

“wherein the disambiguation components present the alternatives to the user in a visual form and allow the user to select from among the alternatives using a voice input” of claim 5,

“where the interaction comprises the concurrent use of said visual mode and said voice mode” of claim 12, and

“wherein the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input” of claim 13.

Furthermore, a secondary consideration of non-obviousness of applicant’s invention is the option provided to the user for selecting the correct uttered word from a plurality of alternate words, if the speech disambiguation system fails to recognize the correct uttered word. Moreover, the speech disambiguation system enables the user to select the correct word in visual mode, voice mode or a combination of voice and visual mode. Since the applicant’s invention offers wider choice and flexibility to the user for selecting the correct uttered word while using a speech disambiguation system, it is more likely to be commercially successful.

Another secondary consideration of non-obviousness of applicant’s invention is that the existing art enables a user to set initial parameters for automatic speech

RECEIVED
CENTRAL FAX CENTER

JUL 21 2008

recognition based on user preferences. However, the existing art fails to provide an option to set parameters based on the end application and requires resetting of the parameters each time based on the application. Hence, there is a long felt but unsolved need for an automatic speech recognition system that is initialized based on the user preferences as well as the end application.

In contrast to the above indicia of non-obviousness, Duan, Lai, and Haddock fail to suggest or implement a method of automatic speech recognition with parameter setting based on user preferences and the end application.

For the reasons stated above, applicant respectfully submits that claims 5, 12, and 13 are not obvious over the cited references, and applicant solicits reconsideration of the rejection and allowance of claims 5, 12, and 13.

Conclusion

Applicant respectfully requests that a timely Notice of Allowance be issued in this case. If, in the opinion of Examiner Rider a telephone conference would expedite the prosecution of this application, Examiner Rider is requested to call the undersigned.

Respectfully submitted,

Date: July 21, 2008

Ashok Tankha, Esq.
Attorney For Applicant
Reg. No. 33,802
Phone: 856-266-5145

Correspondence Address
36 Greenleigh Drive
Sewell, NJ 08080
Fax: 856-374-0246